

A Markovian Model for Joint Observations, Bell's Inequality and Hidden States

Ulrich Faigle¹ and Alexander Schönhuth²

¹ Mathematisches Institut

Universität zu Köln

Weyertal 80, 50931 Köln, Germany

faigle@zpr.uni-koeln.de

² Centrum Wiskunde & Informatica

Science Park 123

1098 XG Amsterdam, The Netherlands

A.Schoenhuth@cwi.nl

Abstract. While the standard approach to quantum systems studies length preserving linear transformations of wave functions, the Markov picture focuses on trace preserving operators on the space of Hermitian (self-adjoint) matrices. The Markov approach extends the standard one and provides a refined analysis of measurements and quantum Markov chains. In particular, Bell's inequality becomes structurally clear. It turns out that hidden state models are natural in the Markov context. In particular, a violation of Bell's inequality is seen to be compatible with the existence of hidden states. The Markov model moreover clarifies the role of the "negative probabilities" in Feynman's analysis of the EPR paradox.

Keywords. Bell's inequality, EPR paradox, hidden state, Markovian operator, negative probability, quantum Markov chain, quantum measurement

1 Introduction

Classical models of physical systems allow a clear distinction between the current state of the system and the probability to extract particular information in that state. The distinction becomes less clear in quantum models where the state of a system is given in terms of its probabilistic description *via* a quantum density, *i.e.*, a self-adjoint matrix with trace 1 and nonnegative eigenvalues. This situation leads to counter-intuitive difficulties that have been pointed out by Einstein, Podolsky and Rosen [8]. While Feynman [10] introduced the idea of "negative probabilities" to resolve the EPR paradox theoretically, Bell [5,6] offered an inequality as a test against the objection of Einstein *et al.*. The violation of Bell's inequality as observed experimentally by Aspect *et al.* [2] seems to confirm the paradoxical nature of the quantum world indeed.

We propose here a model that makes it very natural to distinguish between states and observation probabilities in quantum models. The mathematical key in our analysis is the study of general trace preserving linear operators on the space of self-adjoint

matrices rather than length preserving linear operators on the space of wave functions. This approach is very much in line with the classical Markov model for evolution of states in time and thus yields a comprehensive model for quantum Markov chains as a byproduct, which includes the random walk model of Aharonov *et al.* [1,16] as a special case, for example (Section 4). Without going here into details, we mention that our approach also offers an alternative to Temme *et al.* [21] for quantum Metropolis sampling.

The elementary units in our model are Markov densities, *i.e.*, self-adjoint matrices with trace 1. Markov densities do not necessarily correspond to probabilities but are primarily mathematical descriptions of states of preparation of a system. It follows that a quantum measurement is not necessarily observable in a (macro-)statistical sense if the system is in a general Markov state. Hence it is meaningful to ask when a measurement is observable (Section 3). Moreover, the Markov model yields a refined tool to analyze notions of joint observability of not just two but three or more measurements. We show in Section 5.4 that the Markov perspective provides a novel look at Bell's inequality. In particular, it shows that a violation of Bell's inequality is not contradictory to but compatible with hidden state models. In fact, we argue in Section 6 that hidden state models are natural within (classical or quantum) Markov models of measurements.

As was pointed out previously (see [18]), the EPR paradox can be reframed as resulting from EPR's stipulation that all observables associated with the preparation of a system follow a joint probability distribution which reflects a probability distribution over the inherent hidden states. Feynman suggested to resolve the problem by allowing joint probabilities to be negative, which triggers philosophical difficulties with the interpretation of such quantities. Here, we allow systems to be prepared with a clear view on their hidden states without requiring a preparation to be probabilistic. Instead, we insist that meaningful observations be statistically analyzable. So, according to the standard tenet that quantum measurements are associated with probability distributions, hidden states can perhaps not be measured statistically meaningfully. But they can very well be meaningfully encoded in terms of (Markov) densities.

Motivated by Wigner's [22] distribution, already Dirac [7] and Bartlett [4] had suggested "negative probabilities" to arrive at appropriate mathematical models for seemingly paradoxical quantum world situations (see also Section 5.3) and new probabilistic models have been developed for that purpose (see, *e.g.*, Khrennikov [17]). The Markov approach clarifies the issue: Markov densities describe preparation states and not probability distributions. While those state descriptions may involve negative numbers, there are no negative probabilities. All the probability distributions that are associated with observables in the Markov model are classical (nonnegative). There is no need for new probability theories.

2 Preliminaries

\mathbb{R} denotes the scalar field of real numbers and \mathbb{C} the field of complex numbers $z = a + ib$ (where $a, b \in \mathbb{R}$ and $i^2 = -1$). $\mathbb{C}^{m \times n}$ is the (mn) -dimensional vector space of all

$(m \times n)$ -matrices of the form

$$C = A + iB \quad \text{with } A, B \in \mathbb{R}^{m \times n}. \quad (1)$$

$\bar{C} = A - iB$ is the *conjugate* of $C = A + iB$. The transpose $C^* = \bar{C}^T$ of its conjugate \bar{C} is the *adjoint* of C . $\mathbb{C}^{m \times n}$ is a Hilbert space with respect to the *Hermitian inner product*

$$\langle C|D \rangle := \text{tr}(C^* D) = \sum_{i=1}^m \sum_{j=1}^n \bar{c}_{ji} d_{ij},$$

where the c_{ij} and the d_{ij} denote the coefficients of C and D . $\|C\| := \sqrt{\langle C|C \rangle}$ is the *norm* of C . \mathbb{C}^n is short for $\mathbb{C}^{n \times 1}$ and can also be identified with the space of diagonal matrices in $\mathbb{C}^{n \times n}$:

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{C}^n \quad \longleftrightarrow \quad \text{diag}(v_1, v_2, \dots, v_n) = \begin{bmatrix} v_1 & 0 & 0 & \dots \\ 0 & v_2 & 0 & \dots \\ \vdots & & \ddots & \\ 0 & & \dots & v_n \end{bmatrix}$$

Assuming $m = n$, a matrix $C = A + iB$ with the property $C^* = C$ is *self-adjoint* or *Hermitian*, which means that A is *symmetric* (i.e., $A^T = A$) and B *skew-symmetric* (i.e., $B^T = -B$). Let \mathcal{H}_n denote the collection of all Hermitian $(n \times n)$ -matrices. From the general form (1) one recognizes \mathcal{H}_n as a real Hilbert space of dimension $\dim_{\mathbb{R}}(\mathcal{H}_n) = n^2$.

A matrix $Q = [q_{ij}] \in \mathcal{H}_n$ has real eigenvalues λ_i and a corresponding orthonormal set $\{v_1, \dots, v_n\}$ of eigenvectors $v_i \in \mathbb{C}^n$, yielding the *spectral decomposition*

$$Q = \sum_{i=1}^n \lambda_i v_i v_i^* \quad \text{and hence trace} \quad \text{tr}(Q) = \sum_{i=1}^n q_{ii} = \sum_{i=1}^n \lambda_i. \quad (2)$$

$Q \in \mathcal{H}_n$ is said to be *nonnegative* if all eigenvalues of Q are nonnegative, which is equivalent to the property

$$v^* Q v \geq 0 \quad \text{holds for all } v \in \mathbb{C}^n. \quad (3)$$

A vector $v \in \mathbb{C}^n$ gives rise to a nonnegative element $vv^* \in \mathcal{H}_n$ and one has

$$\langle v|v \rangle = \text{tr}(vv^*). \quad (4)$$

In the case $\text{tr}(vv^*) = 1$, the matrix vv^* is thought to represent a *pure state* of an n -dimensional quantum system.

A matrix $U \in \mathbb{C}^{n \times n}$ is *unitary* if the identity matrix I factors into $I = UU^*$, i.e., if the row (or column) vectors of U form an orthonormal basis for \mathbb{C}^n . So also U^* is unitary. For example, an orthonormal basis $\{v_1, \dots, v_n\}$ of eigenvectors relative to $Q \in \mathcal{H}_n$ gives rise to a unitary matrix U^* with columns v_i . Where $\lambda_1, \dots, \lambda_n$ are the corresponding eigenvalues, the linear operator $x \mapsto Qx$ on \mathbb{C}^n is described with respect to the basis U^* via the transformed matrix

$$UQU^* = \text{diag}(\lambda_1, \dots, \lambda_n) \in \mathcal{H}_n. \quad (5)$$

3 Measurements

Recall that a *quantum measurement* with scale Σ is a finite collection $X = \{M_a \mid a \in \Sigma\}$ of matrices $M_a \in \mathbb{C}^{n \times n}$ such that the self-adjoint matrices $X_a = M_a M_a^*$ yield a *POVM* (= positive operator valued measurement, (cf. [3,19])), i.e., the X_a are non-negative and sum up to the identity:

$$I = \sum_{a \in \Sigma} X_a = \sum_{a \in \Sigma} M_a M_a^*.$$

Assume that the measurement X is applied to a system \mathfrak{S} that is in a state of preparation which is described by some $Q \in \mathcal{H}_n$. If Q is a quantum density, then the parameters

$$p_Q(a) = \langle X_a | Q \rangle = \text{tr}(M_a Q M_a^*)$$

are nonnegative real numbers with sum $\sum_{a \in \Sigma} p_Q(a) = \text{tr}(Q) = 1$. In this case, $p_Q(a)$ is interpreted as the probability for the measurement X to produce the value $a \in \Sigma$, which means that X behaves like a random variable that can be analyzed statistically. Note, however, that Q does not necessarily have to be a quantum density for X be a random variable in the sense above. It suffices to have $\text{tr}(Q) = 1$ and $\text{tr}(X^{(a)} Q) \geq 0$ for all $a \in \Sigma$.

Referring to an element $Q \in \mathcal{H}_n$ with $\text{tr}(Q) = 1$ as a *Markov density*, we therefore say that X is (*statistically*) *observable* relative to the Markov density Q if we have

$$\text{tr}(X^{(a)} Q) \geq 0 \quad \forall a \in \Sigma. \quad (6)$$

The notion of Markov densities Q thus extends and refines the measurement model: The system \mathfrak{S} might be in a (Markov) state of preparation that does not allow X to be analyzed statistically. If Q is nonnegative (and hence a quantum density), X is necessarily observable in Q .

3.1 Markov operators

A *Markov operator* (of *Markovian*) is a linear map $\mu : \mathcal{H}_n \rightarrow \mathcal{H}_n$ that maps Markov densities onto Markov densities. In other words: The Markovian μ is a trace preserving linear operator on \mathcal{H}_n . μ is said to be *nonnegative* if it maps quantum densities onto quantum densities (and hence nonnegative matrices onto nonnegative matrices).

Remark 1 In the case of complete positivity (i.e., $(I \otimes \mu)(A) \geq 0$ for any nonnegative $A \in \mathcal{W} \otimes \mathcal{V}$, where \mathcal{W} is an extra system) a *Markovian operator* is a quantum operation (cf. [19]). The quantum operations formalism aims at modeling the dynamics of open quantum systems and quantum noise, thereby also borrowing from the interrelation between classical noise and classical Markov chains. Time-discrete quantum Markovian dynamics have also been described by trace-preserving quantum operations as quantum channels (see, e.g., [23]). The more general Markovian formalism allows us to interpret quantum systems in terms of hidden states (see Section 5) and to summarize ergodic properties conveniently.

Consider, for example, any quantum measurement $X = \{M_a \mid a \in \Sigma\}$. The associated sum operator

$$Q \mapsto \mu_X(Q) := \sum_{a \in \Sigma} M_a Q M_a^*$$

is nonnegative and trace preserving and hence nonnegative Markovian. In the special case of a measurement $X_U = \{U\}$ with "unary scale" Σ (i.e., $|\Sigma| = 1$), U is a unitary matrix with associated Markov operator

$$Q \mapsto \mu_U(Q) := U Q U^*.$$

Any trace preserving operator $\mu : \mathcal{V} \rightarrow \mathcal{V}$ on some subspace $\mathcal{V} \subseteq \mathcal{H}_n$ extends to a trace preserving operator (and hence Markovian) $\bar{\mu} : \mathcal{H}_n \rightarrow \mathcal{H}_n$ in the obvious way. Consider, for example, the space \mathcal{D} of diagonal matrices $D \in \mathcal{H}_n$. A trace preserving operator $\mu : \mathcal{D} \rightarrow \mathcal{D}$ is described by a

$$\mu(\text{diag}(x)) = \text{diag}(M(x)) \quad (x \in \mathbb{R}^n),$$

where $M \in \mathbb{R}^{n \times n}$ is a *Markov matrix* (i.e., M has column sums 1). μ is nonnegative on \mathcal{D} if and only if all coefficients of M are nonnegative scalars, in which case M is the transition matrix of a (classical) homogeneous Markov process.

3.2 Markov measurements

The quantum measurement formalism generalizes naturally. We call a family $X = \{\mu^{(a)} \mid a \in \Sigma\}$ of linear operators $\mu_a : \mathcal{H}_n \rightarrow \mathcal{H}_n$ a *Markov measurement* with (finite) scale Σ if

$$\mu_X := \sum_{a \in \Sigma} \mu_a \quad \text{is a Markov operator.}$$

X is (*statistically*) *observable* relative to the Markov density Q if

$$\text{tr}(\mu_a(Q)) \geq 0 \quad \text{holds for all } a \in \Sigma.$$

Again, observability means that the measurement X behaves like a random process that produces the output $a \in \Sigma$ with probability $p_Q(a) = \text{tr}(\mu_a(Q))$.

Remark 2 When the μ_a are completely positive, the Markov measurement X is a measurement model in the sense of the quantum operations formalism (see [19]).

The application of an observable X moves the system under investigation with probability $p_Q(a)$ from Q into a state of preparation that is described by the Markov density Q_a such that

$$p_Q(a)Q_a = M_a Q M_a^* \quad \text{and hence} \quad p_Q(a) = \text{tr}(M_a Q M_a^*).$$

So the expected Markov density after the application of X is

$$\sum_{a \in \Sigma} p_Q(a) Q_a = \mu_X(Q). \quad (7)$$

Multinomial expansion of the Markov measurement $X = \{\mu^{(a)} \mid a \in \Sigma\}$ yields

$$(\mu_X)^t = \left(\sum_{a \in \Sigma} \mu_a \right)^t = \sum_{a_1, \dots, a_t \in \Sigma} \mu_{a_t} \mu_{a_{t-1}} \dots \mu_{a_1} =: \sum_{w \in \Sigma^t} \mu_w$$

and exhibits $X^t := \{\mu_w \mid w \in \Sigma^t\}$ as a Markovian measurement with scale Σ^t for any $t = 1, 2, \dots$. We say that X is t -observable in Q if X^t is observable in Q . Obviously, X is always 0-observable.

Lemma 1. *Assume that the Markov measurement X is t -observable in Q . Then X is k -observable in Q for all $k \leq t - 1$.*

Proof. It suffices to consider $k = t - 1$. Let $w = a_1 \dots a_{t-1} \in \Sigma^{t-1}$ be arbitrary. Since X^t is observable and μ_X Markovian, one has

$$\text{tr}(\mu_w(Q)) = \text{tr}(\mu_X(\mu_w(Q))) = \sum_{a \in \Sigma} \text{tr}(\mu_{wa}(Q)) \geq 0.$$

◊

If X is k -observable in Q , the application of X^k yields the expected Markov density $\mu_X^k(Q)$. The following is straightforward.

Lemma 2. *Assume that the Markov measurement X is t -observable in Q . Then X is observable in $\mu^k(Q)$ for all $k \leq t - 1$.*

◊

We say that X is *completely observable* in Q if X is t -observable in Q for all $t \geq 1$.

Remark 3 *With the methods of [9] one can show that completely observable Markov measurements X yield asymptotically mean stationary processes. Such processes are known to have good ergodic properties (see [13, 12]). For example, the conditional entropies converge to a limit*

$$H_\infty(X) = \lim_{t \rightarrow \infty} \frac{1}{t} H(X^t) = \lim_{t \rightarrow \infty} H(X^t | X^{t-1}).$$

In the case of diagonal matrices, complete observability is equivalent to the much studied concept of finite-dimensional (classical) stochastic processes or observable operator models that provide a proper generalization of classical (possibly hidden) discrete Markov chains with finite alphabets (see [11, 14, 15]). We are lead to the latter model in more detail in Section 6.

4 Markov chains

Let μ be a Markov operator on \mathcal{H}_n . Then any Markov density P gives rise to a chain of Markov densities $\mu^t(P)$,

$$(\mu, P) := \{\mu^t(P) \mid t = 0, 1, \dots\},$$

where μ^0 is the identity operator on \mathcal{H}_n . We refer to μ as the *evolution operator* of the *Markov chain* (μ, P) . In the case $\mu(P) = P$, the chain (μ, P) is *stationary*. Note that classical Markov chains correspond to Markov chains of nonnegative diagonal matrices in the general model.

Remark 4 *The Markovian evolution operator μ does not have any density or probability interpretation in its own right but simply describes how the densities in a chain (μ, P) evolve over time.*

The Markov chain (μ, P) is *bounded* if there exists some $c \in \mathbb{R}$ such that

$$\langle \mu^t(P) | \mu^t(P) \rangle = \text{tr}(\mu^t(P)^2) \leq c \quad \text{holds for all } t.$$

Assume, for example, that μ is nonnegative and P a quantum density. Then the Markov chain (μ, P) is bounded because each $\mu^t(P)$ is a quantum density and therefore satisfies $\text{tr}(\mu^t(P)^2) \leq 1$.

Evolution of wave functions. Any unitary matrix $U \in \mathbb{C}^{n \times n}$ induces a nonnegative Markovian $\mu_U(C) = UCU^*$. On the other hand, a wave function $v \in \mathbb{C}^n$ of norm $\|v\| = 1$ is transformed under U into the wave function Uv of norm $\|Uv\| = 1$, which corresponds to the pure quantum state

$$(Uv)(Uv)^* = Uvv^*U^* = \mu_U(vv^*).$$

So (μ_U, vv^*) is the Markov chain of the pure quantum states that correspond to the wave function evolution $v \mapsto U^t v$ for $t = 0, 1, \dots$

Quantum random walks on graphs. A model for a quantum random walk on a directed regular graph G has been proposed by Aharonov *et al.* [1] (see also Kempe [16]). A natural generalization of that model to arbitrary directed graphs on V is the following.

Given the graph $G = (V, \mathcal{E})$ with node set V and edge set $\mathcal{E} \subseteq V \times V$, we identify the elements $e \in \mathcal{E}$ with the elements in the standard basis of unit vectors in \mathbb{C}^n . A quantum walk on G is now described in terms of a Markovian operator $\mu : \mathcal{H}_n \rightarrow \mathcal{H}_n$. The walk on G starts at time $t = 0$ in a quantum density state $P^{(0)}$ and consists in subsequent density changes according to

$$P^{(t)} := \mu(P^{(t-1)}) \quad \text{for all } t = 1, 2, \dots$$

The quantum walk corresponds to the Markov chain $(\mu, P^{(0)})$ in our terminology. Let us define for every node $v \in V$ the diagonal matrix $X_v = \text{diag}(x^{(v)}) \in \mathcal{H}_n$, where the vector $x^{(v)} \in \mathbb{C}^n$ has the components

$$x_e^{(v)} := \begin{cases} 1 & \text{if } e \in \mathcal{E} \text{ is of the form } e = (v, w) \\ 0 & \text{otherwise.} \end{cases}$$

Note that the family $X = \{X_v \mid v \in V\}$ constitutes a POVM. If all $P^{(t)}$ are nonnegative, the application of the matrices $X^{(v)}$ yields the nonnegative parameters

$$p_t(v) = \langle X_v | P^{(t)} \rangle = \text{tr}(X_v P^{(t)}) \geq 0,$$

which means that X is observable. $p_t(v)$ is then interpreted as the probability for the random walk to be in node v at time t .

Aharonov *et al.* [1] assume a unitary wave function evolution and hence a non-negative Markovian operator of the type $\mu_U(P) = UPU^*$. Using the assumed regularity of the underlying graph G , they derive an ergodic property of their quantum walk and show that the following limit of averages exists in their model:

$$\bar{p}(v) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=0}^{t-1} p_k(v).$$

We will exhibit a general convergence property for bounded Markov chains below.

Remark 5 *Our general quantum random walks need not respect the structure of the underlying graph. This is purely for formal convenience: Our model incorporates the most general graph-respecting concepts of quantum walks suggested in [1].*

It is of interest to know whether a Markov chain converges towards a stationary behavior. We consider a fixed Markovian $\mu : \mathcal{H}_n \rightarrow \mathcal{H}_n$.

Theorem 1. *Let (μ, P) be a bounded Markov chain. Then the limit of averages*

$$\tilde{P} = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t \mu^k(P) \quad (8)$$

exists and (μ, \tilde{P}) is a stationary Markov chain. Moreover, if μ is nonnegative and P a quantum density, then also \tilde{P} is a quantum density.

If the limit (8) exists, it is clear that $\text{tr}(\tilde{P}) = \text{tr}(P) = 1$ holds and (μ, \tilde{P}) is stationary. Moreover, if μ preserves quantum densities, then each $\mu^t(P)$, and therefore each average, is a quantum density. So it remains to prove the existence of \tilde{P} . It is convenient to base the proof on the following lemma.

Lemma 3 ([9]). *Let V be a finite-dimensional normed vector space over \mathbb{C} and consider the linear operator $F : V \rightarrow V$. The following statements are equivalent:*

- (a) $\bar{v} = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^{t-1} F^k(v)$ exists for all $v \in V$.
- (b) For every $v \in V$, there exists some finite bound $c \in \mathbb{R}$ such that $\|F^t(v)\| \leq c$ holds for all $t \geq 0$.

We want to apply Lemma 3 to the (complex) subspace $V \subseteq \mathbb{C}^{n \times n}$ that is generated by (μ, P) with the norm $\|C\| = \sqrt{\text{tr}(C^*C)}$. To this end, we augment P to a basis $\{Q_0, Q_1, \dots, Q_m\} \subseteq (\mu, P)$ for V with $Q_0 = P$ and define the linear operator $F : V \rightarrow V$ via

$$F\left(\sum_{j=0}^m r_j Q_j\right) := \sum_{j=0}^m r_j \mu(Q_j).$$

Then F agrees with μ on the Markov chain (μ, P) . Let c_1 be a bound on (μ, P) and observe from the triangle inequality:

$$\|F^t\left(\sum_{i=1}^m r_i Q_i\right)\| \leq \sum_{i=1}^m |r_i| \cdot \|\mu^t(Q_i)\| \leq \sum_{i=1}^m |r_i| c_1 =: c.$$

So F satisfies condition (b) and hence also (a) of Lemma 3, which establishes the convergence of the averages in Theorem 1 with the choice $v = P$.

Corollary 1. *Let (μ, P) be a bounded Markov chain and $X : \mathcal{H}_n \rightarrow \mathbb{R}$ any linear functional. Then*

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t X(\mu^k(P)) = X(\tilde{P}).$$

◇

Remark 6 *In contrast to density averages, wave function averages in a wave function evolution will generally not converge to a wave function. If U is a unitary matrix that does not have 1 as an eigenvalue, one has*

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t U^k v = 0 \quad \text{for all } v \in \mathbb{C}^n.$$

5 Hidden states and Bell's inequality

Let \mathfrak{S} be a physical system with a finite set $\Omega = \{\omega_1, \dots, \omega_N\}$ of *hidden states* and assume that \mathfrak{S} is (definitely) in one of the N possible hidden states $\omega \in \Omega$ at any discrete time $t = 0, 1, \dots$. It may not be possible to observe the true hidden state $\omega = \omega^{(t)}$ at time t directly. So we want to collect statistical information on the true state.

5.1 Information functions

An *information function* with scale Σ is a function $X : \Omega \rightarrow \Sigma$. Since Ω is finite, we may assume Σ to be finite as well. Equivalently, X can be viewed as a POVM $\{X_a \mid a \in \Sigma\}$ of diagonal matrices $X_a = \text{diag}(x^{(a)})$, with $x_\omega^{(a)} = 1$ if $X(\omega) = a$.

Similar to our terminology in Section 3, we call a vector $q \in \mathbb{R}^\Omega$ with components q_ω a *Markov state* of \mathfrak{S} if

$$\sum_{\omega \in \Omega} q_\omega = 1.$$

For any $a \in \Sigma$, we set

$$p_q(a) := \sum_{\omega: X(\omega)=a} q_\omega$$

and say that the information function X is (*statistically*) *observable* in q if all the $p_q(a)$ are nonnegative numbers. In the case of observability and a real-valued information function X , we obtain the well-defined expectation

$$E_q(X) := \sum_{x \in \Sigma} x \cdot p_q(x) = \sum_{\omega \in \Omega} X(\omega) q_\omega.$$

5.2 Joint observations

We say that the k information functions $X_1 : \Omega \rightarrow \Sigma_1, \dots, X_k : \Omega \rightarrow \Sigma_k$ on the system \mathfrak{S} are *jointly observable* in the Markov state q if the composite information function $X : \Omega \rightarrow \Sigma$ with

$$X(\omega) := (X_1(\omega), \dots, X_k(\omega)) \quad \text{and} \quad \Sigma := \Sigma_1 \times \dots \times \Sigma_k$$

is observable in q . The following statement is easy to verify.

Lemma 4. *Assume that the collection of k information functions X_1, \dots, X_k is jointly observable in the Markov state q , then every subcollection X_{i_1}, \dots, X_{i_m} is jointly observable in q . In particular, every individual information function X_i is observable. Moreover, if the X_i are real-valued, also every product $X_i X_j$ is observable in q .*

◇

Hence, if two information functions X and Y on the system \mathfrak{S} are real-valued and jointly observable, their product XY is statistically observable and has a well-defined expectation $E(XY)$.

Clearly, any collection of information functions is jointly observable in any non-negative Markov state q , *i.e.*, classical probability distribution q on Ω .

5.3 Feynman's approach to the EPR paradox

Feynman [10] has given a mathematical model to explain the Einstein, Rosen and Podolsky (EPR) paradox (see also Scully *et al.* [20]). In essence, Feynman provides the example of two information functions X, Z for the spin along the $+x$ and $+z$ axis of a spin 1/2 system \mathfrak{S} with $\Omega = \{(++), (+-), (-+), (--) \}$ as in the following table:

	(++)	(+-)	(-+)	(--)
X	+	+	-	-
Z	+	-	+	-

and the "joint probability distribution":

$$\begin{aligned} P(++) &= [1 + \langle \hat{\sigma}_z \rangle + \langle \hat{\sigma}_x \rangle + \langle \hat{\sigma}_y \rangle]/4 \\ P(+-) &= [1 + \langle \hat{\sigma}_z \rangle - \langle \hat{\sigma}_x \rangle - \langle \hat{\sigma}_y \rangle]/4 \\ P(-+) &= [1 + \langle \hat{\sigma}_z \rangle + \langle \hat{\sigma}_x \rangle - \langle \hat{\sigma}_y \rangle]/4 \\ P(-- &= [1 - \langle \hat{\sigma}_z \rangle - \langle \hat{\sigma}_x \rangle - \langle \hat{\sigma}_y \rangle]/4, \end{aligned}$$

where $\langle \hat{\sigma}_x \rangle, \langle \hat{\sigma}_y \rangle, \langle \hat{\sigma}_z \rangle$ are the Pauli spin operators. Noticing that those parameters $P(xz)$ might take on negative values, Feynman interprets them as (possibly) *negative* probabilities and argues that probability theory should be expanded in this direction.

In our terminology, the $P(xz)$ yield a blueprint for attainable Markov states (depending on the actual values of the spin operators), where X and Z are not always guaranteed to be jointly statistically observable. For example, the situation

$$\langle \hat{\sigma}_x \rangle = \langle \hat{\sigma}_y \rangle = \langle \hat{\sigma}_z \rangle = 1/2,$$

would result in the Markov state $q = (5/8, 1/8, 3/8, -1/8)$, where X and Z are individually but not jointly observable.

5.4 Bell's inequality

The well-known inequality of Bell [5,6] takes the form (9) in our context as a statement on the expectations of products of pairs of information functions.

Lemma 5 (Bell's inequality). *Let $X, Y, Z : \Omega \rightarrow \{-1, +1\}$ be arbitrary information functions on the system \mathfrak{S} . If X, Y and Z are jointly observable in the Markov state q , then the following inequality holds:*

$$|E_q(XY) - E_q(YZ)| \leq 1 - E_q(XZ). \quad (9)$$

Proof. Any choice of $x, y, z \in \{-1, +1\}$ satisfies the inequality $|xy - yz| \leq 1 - xz$. Because of the joint observability assumption, all the observation probabilities

$$p_q(x, y, z) = \Pr\{X = x, Y = y, Z = z\}$$

are nonnegative real numbers that sum up to 1. So we conclude

$$\begin{aligned} |E_q(XY) - E_q(YZ)| &= \left| \sum_{x,y,z} (xy - yz)p_q(x, y, z) \right| \leq \sum_{x,y,z} |xy - yz|p_q(x, y, z) \\ &\leq \sum_{x,y,z} (1 - xz)p_q(x, y, z) = 1 - E_q(XZ). \end{aligned}$$

◊

Bell's inequality may be violated by information functions that are pairwise but not jointly observable. Consider a system \mathfrak{S} with a set $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$ of five

hidden states, for example, and three information functions $X, Y, Z : \Omega \rightarrow \{-1, +1\}$ as in the following table:

	ω_1	ω_2	ω_3	ω_4	ω_5
X	-1	+1	-1	-1	-1
Y	+1	+1	-1	+1	-1
Z	+1	+1	+1	-1	-1

One can check that X, Y, Z are pairwise observable in the Markov state

$$q = (-1/3, 1/3, 1/3, 1/3, 1/3)$$

and yield the product expectations

$$E_q(XY) = +1, E_q(YZ) = -1/3, E_q(XZ) = +1,$$

which violate Bell's inequality (9).

Remark 7 *Experimental results seem to indicate that quantum systems may violate Bell's inequality (see, e.g., Aspect et al. [2]). This is sometimes interpreted as showing that quantum mechanics does not admit a theory with hidden variables. The Markovian picture makes it clear that a violation of Bell's inequality only shows that the system is studied in terms of measurements that are perhaps pairwise but not jointly observable. The existence of definite but hidden states is not excluded. In fact, an experimentally observed violation of Bell's inequality suggests that one should not place a priori non-negativity restrictions on concepts of states into which a system can be prepared.*

6 Measurement processes and hidden states

Let us now argue that statistical models for the analysis of measurements involve Markov states with possibly negative components quite naturally.

Suppose that some measurement X with finite scale Σ is made on some (not further specified) system \mathfrak{S} at discrete times $t = 1, 2, \dots$. Let X_t be the result observed at time t and assume that the results of the measurements are jointly observable in the sense that the X_t are random variables with joint probabilities

$$p(a_1 \dots a_t) = \Pr\{X_1 = a_1, \dots, X_t = a_t\} \quad \text{for all } a_1, \dots, a_t \in \Sigma.$$

For any $w = a_1 \dots a_t$ and $v = a_{t+1} \dots a_{t+s}$, the conditional probability

$$p(v|w) = \Pr\{X_{t+1} = a_{t+1}, \dots, X_{t+s} = a_{t+s} | X_1 = a_1, \dots, X_t = a_t\}$$

is the probability for correctly predicting the sequence v to occur in the next s observations when w is obtained up to time t . (For formal reasons, it is convenient to assume that the *empty symbol* \square is issued at time $t = 0$ with probability $p(\square) = 1$.) We collect all the prediction probabilities into the (infinite) *prediction matrix*

$$\mathcal{P} = [p(v|w)].$$

\mathcal{P} is a Hankel matrix. The rank $\text{rk}(\mathcal{P})$ is also known as the *dimension* of X . For example, classical (possibly hidden) Markov chains with finite alphabet Σ can be shown to be finite-dimensional ([11,14,15]). Moreover, finite-dimensional processes are asymptotically mean stationary (see [9]).

In the case $m = \text{rk}(\mathcal{P}) < \infty$ it is not difficult to see (cf. [15]) that (X_t) admits an *observable operator model* in the following sense:

- There is a family $X = \{M_a | a \in \Sigma\}$ of real $(m \times m)$ -matrices M_a and a Markov state $\pi \in \mathbb{R}^m$ such that for all $a_1, \dots, a_t \in \Sigma$
 - (1) $p(a_1 \dots a_t) = \text{tr}(\text{diag}(M_{a_t} M_{a_{t-1}} \dots M_{a_1} \pi))$;
 - (2) $M = \sum_{a \in \Sigma} M_a$ is a Markov matrix.

In our terminology, this means that X represents a Markov measurement with respect to diagonal matrices that is completely observable relative to the Markov density $\Pi = \text{diag}(\pi)$.

Remark 8 *Finite-dimensional measurement processes that do not admit an observable operator model with only nonnegative Markov states are known to exist* (cf. [15]).

Let (X_t) be a measurement process on a system \mathfrak{S} with finite alphabet Σ and finite dimension m and the associated observable operator model $(\{M_a | a \in \Sigma\}, \pi)$. Consider

$$\Omega = \{(a, j) | a \in \Sigma, j = 1, \dots, m\}$$

as the set of $N = |\Sigma| \cdot m$ hidden states of \mathfrak{S} with the information function

$$X : \Omega \rightarrow \Sigma \quad \text{such that} \quad X(a, j) = a.$$

With each $M_a = [m_{ij}^{(a)}] \in \mathbb{R}^{m \times m}$, we associate the $(N \times N)$ -matrix

$$\overline{M}_a = [m_{(b,i),(b,j)}] \quad \text{with} \quad m_{(b,i),(c,j)} = \begin{cases} m_{ij}^{(a)} & \text{if } b = a \\ 0 & \text{if } b \neq a. \end{cases}$$

Let furthermore $\overline{\pi}$ be the N -dimensional vector with coordinates

$$\overline{\pi}(a, j) = \pi(j)/|\Sigma|.$$

By construction, we have now for all $w = a_1 \dots a_t$,

$$p(w) = \text{tr}(\text{diag}(\overline{M}_w \overline{\pi})).$$

The process (X_t) can thus be interpreted as emanating from \mathfrak{S} via the information function X . The system is prepared to be in the Markov state $\overline{\pi}$ at time $t = 0$. At time $t \geq 1$, the system is in a hidden state $\omega \in \Omega$ yielding the information $X(\omega) = a$ with probability

$$\Pr\{X_t = a\} = \sum_{\omega: X(\omega)=a} \overline{\pi}^{(t)}(\omega) \quad (\text{where } \overline{\pi}^{(t)} = \overline{M}^t \overline{\pi}).$$

References

1. D. AHARONOV, A. AMBAINIS, J. KEMPE, U. VAZIRANI: *Quantum walks on graphs*, Proc. 33th STOC, 60-69, ACM, New York, NY, 2001.
2. A. ASPECT, J. DALIBARD, G. ROGER: *Experimental tests of Bell's inequalities using time-varying analyzers*, Phys. Rev. Lett. 49, 1804 (1982).
3. O.E. BARNDORFF-NIELSEN, R.D. GILL, P.E. JUPP, *On quantum statistical inference*, J. Roy. Statist. Soc. B (2003), 775-816.
4. M. BARTLETT, *Negative probability*, Math. Proc. Camb. Phil. Soc. 41, 71-73 (1945).
5. J.S. BELL: *On the Einstein Podolsky Rosen paradox*, Physics 1, 195-200 (1964).
6. J.S. BELL: *On the problem of hidden variables in quantum mechanics*, Rev. Mod. Phys. 38, 447-452 (1966).
7. P.A.M. DIRAC: *The physical interpretation of quantum mechanics*, Proc. Royal Soc. London A 180, 1-39 (1942).
8. A. EINSTEIN, B. PODOLSKY, N. ROSEN: *Can quantum mechanical descriptions of physical reality be considered complete?*, Phys. Rev. 47, 777-780 (1935).
9. U. FAIGLE AND A. SCHÖNHUTH: *Asymptotic mean stationarity of sources with finite evolution dimension*, IEEE Trans. Information Theory 53 (2007), 2342-2348.
10. R.P. FEYNMAN: *Quantum Implications, Essays in Honour of David Bohm*, B.J. Hiley and F.D. Peat eds., Routledge and Kegan Paul, London, 235-246 (1987).
11. E.J. GILBERT: *On the identifiability problem for functions of finite Markov chains*, Ann. Math. Stat. 30, 688-697, 1959.
12. R.M. GRAY, *Probability, Random Processes, and Ergodic Properties*, Springer-Verlag, New York, 1988.
13. R.M. GRAY, *Entropy and Information Theory*, Springer-Verlag, New York, 1990.
14. H. ITO, S.-I. AMARI, K. KOBAYASHI: *Identifiability of hidden Markov information sources and their minimum degrees of freedom*, IEEE Transactions on Information Theory 38, 324-333 (1992).
15. H. JAEGER: *Observable operator models for discrete stochastic time series*, Neural Computing 12, 1371-1398 (2000).
16. J. KEMPE: *Quantum random walks: an introductory overview*, Contemporary Physics 44, 307-327 (2003).
17. A. KHRENNIKOV: *Interpretations of Probability*, VSP Science, Zeist, 1999.
18. W.M. DE MUYNCK AND OMAR ABU-ZEID, *On an alternative interpretation of the Bell inequalities*, Phys. Lett. 100A, 485-489 (1984).
19. M. NIELSEN AND I. CHUANG, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.
20. M.O. SCULLY, H. WALTHE AND W. SCHLEICH, *Feynman's approach to negative probability in quantum mechanics*, Phys. Rev. A 49, 1562-1566 (1994).
21. K. TEMME, T.J. OSBORNE, K. VOLLBRECHT AND F. VERSTRAETE, *Quantum Metropolis Sampling*, arXiv.0911.3635v2 [quant-ph] (2010).
22. E.P. WIGNER, *On the quantum correction for thermodynamic equilibrium*, Phys. Rev. 40, 749-759 (1932).
23. M.M. WILDE, *From Classical to Quantum Shannon Theory*, arXiv:1106.1445 (2011).